DATA400 – Idea 1 – Kevin Tran

1. General ideas

- *The* **Question**: Does a higher national GDP and "developed" status, such as that of the United States, necessarily result in a better patient experience compared to select countries in Asia?
- I'm thinking of comparing the rating, the costs of hospitals in East & West to see if the difference in GDP shows the difference in quality of healthcare (price, efficiency, waiting time…). To do that, I'm thinking of gathering reviews of hospitals.
- My sequence is in ECON so I think this might be relatable.

2. Detailed

- This project plans to employ a multiple regression analysis to model the relationship between economic factors and healthcare quality.
    - Healthcare Quality Data (Dependent Variables):
        - Data will be gathered by scraping public reviews from sources like Google Maps and other hospital review aggregators.
        - Key metrics will include quantitative star ratings and qualitative data from text reviews, which can be analyzed for mentions of waiting times, costs, and care efficiency.
    - Economic Data (Independent Variables):
        - To address the feedback on data sourcing, economic data for the hospital locations (e.g., by county or ZIP code) will be obtained from official sources like the U.S. Census Bureau or the Bureau of Labor Statistics.
        - Specific indicators will include median household income, poverty rates, and percentage of uninsured residents.
- I think this will be of great use because in the past, I have seen many cases that many patients need to wait for doctors from another country to come over. Or even a family bring their child/their parents to the West, only to realize that they will be waiting for a doctor from their home country to come over and treat them. This will probably create some insights on where should a middle-income family choose for treatment base on their urgency. This might also reveal the problem with the healthcare systems of certain countries, and probably propose ideas to improve the conditions in developing countries.
- Stakes: The greatest risk is oversimplification and cultural stereotyping. Presenting findings as "U.S. healthcare is better/worse than Asian healthcare" would be a dangerous generalization. Such a conclusion could promote bias, misinform patients seeking care abroad, and unfairly criticize national healthcare workforces.
- Ethical Implications:
    - Cultural Bias: The interpretation of "good" service can vary across cultures. A direct, critical review might be common in the U.S. but considered rude and

thus rare in other cultures, skewing the data. The analytical model must attempt to account for this cultural reporting bias.

- Data Representation: Internet access and the practice of leaving online reviews may not be as widespread or representative across all demographics in the selected Asian countries compared to the U.S. This could lead to a dataset that over-represents urban, wealthier, and younger populations.
- Complex Causality: It is crucial to emphasize that any correlation found between GDP and patient satisfaction does not imply causation. A country's healthcare quality is a product of its history, political system, culture, and infrastructure-factors that cannot be fully recorded by economic data alone.

3. Extra

- For additional idea development, I'm thinking of also scraping reviews from e.g Reddit, Quora, … and classifying them through a ML model. However this might be a bit complicated so I'm still considering if that is viable for the 2 week period.